ADVANCES IN MATHEMATICS 30, 171–201 (1978)

A Theory of Entropy in Fourier Analysis*

ROBERT A. FEFFERMAN

Department of Mathematics, University of Chicago, Chicago, Illinois 60637

This thesis deals with a certain set function called entropy and its applications to some problems in classical Fourier analysis. For a set $S \subsetneq [0, 1/e]$ the entropy of S is defined by $E(S) = \inf_{S \subset \bigcup_k I_k,\, I_k \text{ intervals}} \sum_k |I_k| \log(1/|I_k|)$. We begin by using notions related to entropy in order to investigate the maximal operator M_Ω given by $M_\Omega(f)(x) = \sup_{r>0}(1/r^n) \int_{|t| \leqslant r} \Omega(t)\, |f(x+t)|\, dt$, $f \in L^1(R^n)$, where Ω is a positive function, homogeneous of degree 0, and satisfying a certain weak smoothness condition. Then the set function entropy is investigated, and certain of its properties are derived. We then apply these to solve various problems in differentiation theory and the theory of singular integrals, deriving in the process, entropic versions of the theorems of Hardy and Littlewood and Calderón and Zygmund.

INTRODUCTION

This paper deals with a certain set function, entropy, and its applications to problems in classical analysis. We begin by considering a nonisotropic maximal function, M_Ω, defined as

$$M_\Omega(f)(x) = \sup_{r>0} \frac{1}{r^n} \int_{|t| \leqslant r} \Omega(t)\, |f(x+t)|\, dt, \qquad x \in R^n$$

(here f is a locally integrable function on R^n, and Ω is a positive function homogeneous of degree 0). This problem was posed by Stein [8]. It will be shown that under certain weak restrictions on the kernel Ω, the operator M_Ω is of weak type (1, 1); that is,

$$m\{M_\Omega(f) > \alpha\} \leqslant (C_\Omega/\alpha)\, \|f\|_1,$$

and the point is that the crucial quantity which controls M_Ω is something which is intimately connected with the concept of entropy. (At this point we are motivated by the work of C. Calderón on differentiation with respect to starlike sets, where a set function similar to entropy is introduced (see [4]).) This

* A dissertation presented to the faculty of Princeton University in candidacy for the degree of doctor of philosophy.

0001-8708/78/0303–0171$05.00/0

discussion motivates the consideration of entropy, but there are other reasons for studying this quantity. So after we have discussed the operator M_Ω and the corresponding theory for singular integrals, we shall look at entropy as an entity in itself, without worrying about particular applications. On the interval $[0, 1/e] = I^*$ the entropy $E(S)$ of a set $S \subset [0, 1/e]$

$$E(S) = \inf_{S \subset \cup I_k} \sum_k |I_k| \log \frac{1}{|I_k|},$$

where here the inf is taken over all sequences of *intervals* $I_k \subset [0, 1/e]$ which cover S. The properties of E are discussed in some detail. For example, it is shown that if we extend the notion of entropy so that we can speak about the entropy of a function (the entropy, $\mathfrak{J}(f)$, of a function f will, roughly speaking, be the integral of f with respect to the set function E) then entropy of functions is a normed quantity.

Now, at this point, it is important to have in mind the main idea of what it is that entropy measures. First, if S is a subset of I^*, then, since in the definition of $E(S)$ we must cover S by a collection of intervals, the Lebesgue measure of S enters into the computation of the entropy. On the other hand in the definition of $E(S)$, we must not only cover S by intervals I_k, but cover S as to minimize the sum

$$\sum |I_k| \log(1/|I_k|).$$

In an arbitrary cover $\{I_k\}$ of S it may well happen that there are two disjoint intervals I_j and I_k with the property that if I is the smallest interval containing both I_j and I_k then

$$|I| \log \frac{1}{|I|} \leqslant |I_j| \log \frac{1}{|I_j|} + |I_k| \log \frac{1}{|I_k|}$$

and yet $|I| > |I_j| + |I_k|$. Thus, minimizing $\sum |I_k| \log(1/|I_k|)$ involves a delicate interplay between the Lebesgue measure of S and the geometry of S. Unless $|S|$ is small, the entropy cannot be small. On the other hand if $|S|$ is fixed, the quantity $E(S)$ is determined by how spread out the mass of S is.

What we are saying, then, is that entropy (of, say, a function) cares not only about how large a function is but also about the *geometry* of the set where the function is large.

It is with this point of view that we shall proceed, studying the behavior of the entropy of functions under the action of familiar operators, such as the Hardy–Littlewood Maximal operator, and singular integrals of the Calderón–Zygmund variety. And each estimate that we obtain for entropy will have as immediate corollaries estimates in a more familiar setting.

These inequalities, for example, give upper bounds for the distribution function of the Hilbert transform of a function which is very slightly smooth.

And the results we obtain also apply to give results in the area of exceptional sets. This last statement is not so surprising when we observe that the definition of Hausdorff measure resembles that of entropy. Actually, this difference (the fact that to get the Hausdorff measure, H, of S with respect to $\beta(t) = t \log(1/t)$, we must put restrictions on the diameters of the intervals covering S, whereas in the case of entropy, no such restriction occurs) is a fundamental one. Hausdorff measure is a Borel measure while entropy, by its very nature, is highly non-additive. Although from a certain point of view, this makes entropy harder to deal with, the crucial advantage of entropy over Hausdorff measure is that while the entropy and H vanish for exactly the same sets, certain quantitative estimates for operators could never be made with H-measure and must therefore be approached only via entropy. More specificly, if f is a continuous nonzero function, then $H\{|f| > \alpha\} = \infty$, for each $\alpha > 0$, while if $\lambda(\alpha) = E\{|f| > \alpha\}$ (λ is the entropic distribution function of f) then λ is an interesting quantity subject to some sort of estimation.

A comment about the nature of our estimates is in order. It was proved by Carleson [5] that the Hardy–Littlewood maximal operator, M, satisfies the L^2 estimate

$$C_K\{x \mid M(f)(x) > \alpha\} \leqslant \frac{A}{\alpha^2} \left(\iint_{T^2} \frac{|f(x) - f(y)|^2}{|x - y|} \, dx \, dy + \|f\|_2^2 \right)$$

where C_k is the capacity with respect to the kernel

$$K(x) = 1 \Big/ |x| \left(\log \frac{1}{|x|} \right)^2,$$

and T^2 is $[0, 2\pi] \times [0, 2\pi]$, the torus. Someqhat later Adams [1] was able to extend this result to a capacitory weak type (p, p) inequality for $1 < p < \infty$. What we shall do is produce the corresponding L^1 estimate

$$E\{x \mid M(f)(x) > \alpha\} \leqslant \frac{A}{\alpha} \left(\iint_{T^2} \frac{|f(x) - f(y)|}{|x - y|} \, dx \, dy + \|f\|_1 \right). \qquad (*)$$

As for Calderón–Zygmund singular integral operators T we shall obtain an entropic estimate which gives, in particular, the weaker capacitary estimate

$$C_K\{x \mid |Tf(x)| > \alpha\} \leqslant \frac{A_T}{\alpha} \left(\|f\|_1 + \iint_{T^2} \frac{|f(x) - f(y)|}{|x - y|} \, dx \, dy \right).$$

quantity in parenthesis on the right side of the preceding equation $\|f\|_{D^1}$.) Some other applications of various sorts are given in the paper for the concept of entropy. It is an easy matter to show, for example, that for $f \in L^1(R^3)$ satisfying a certain weak smoothness condition similar to the Dini norm on the right-hand side of (*) we have

$$\lim_{s,t\to 0} \frac{1}{8s^2t^2} \int_{z-st}^{z+st} \int_{y-t}^{y+t} \int_{x-s}^{y+s} f(\xi, \zeta, \eta)\, d\xi\, d\zeta\, d\eta = f(x, y, z)$$

for a.e. $(x, y, z) \in R^3$. Yet another example of an application of entropy to Fourier analysis is a real variable proof that any $f \in L^1(T)$ satisfying the L^1 Dini condition $\|f\|_{D^1} < \infty$ belongs to the class $L(\log^+ L)$. This fact is useful in the theory of singular integrals.

It also seems likely that enctropy has an intimate connection to the pointwise convergence of Fourier series. For the corresponding capacitory theory the reader should consult [13] for the theorem of Salem and Zygmund.

Thus, we see that entropy has a number of applications to classical Fourier analysis. And because of the simultaneous affinity of this quantity to measures of exceptional sets, to the class $L(\log L)$, to the method of rotations, and to the basic geometry of sets, it seems likely that the applications given here are but a very small sample of those possible.

1. A Nonisotropic Maximal Operator

In this section, we begin by posing a question having to do with a certain maximal operator. More specificially, let Ω be a positive function defined on R^n, homogeneous of degree 0, and satisfying $\int_{S^{n-1}} \Omega(x)\, d\sigma(x) = 1$ (here $d\sigma$ is the measure of surface area on the unit sphere S^{n-1} in R^n.) Let us define the operator M_Ω by

$$M_\Omega(f)(x) = \sup_{r>0} \frac{1}{r^n} \int_{|t|\leqslant r} \Omega(t)\, |f|\, (x+t)\, dt,$$

$f \in L^1(R^n)$. Then M_Ω is the sup of weighted averages of $|f|$ over balls, where in the weighted averages we "count" different directions $t \in S^{n-1}$ differently according as whether $\Omega(t)$ is large or small. So M_Ω is a nonisotropic version of the Hardy–Littlewood maximal operator, which is bounded on L^p for $1 < p \leqslant \infty$ and is of weak type 1–1. The question which concerns us here is "What are the boundedness properties of M_Ω on the various L^p spaces?"

The way one can handle the operator M_Ω is to dominate it by an operator M_Ω^* which is a weighted average of one-dimensional Hardy–Littlewood maximal operators. This is the so-called "method of rotations" of Calderón and Zygmund in [3]. We have

$$M_\Omega(f) \leqslant M_\Omega^*(f) = \int_{S^{n-1}} \Omega(t)\, M^t(f)\, d\sigma(t),$$

where for $t \in S^{n-1}$

$$M^t(f)(x) = \sup_{s>0} \frac{1}{2s} \int_{-s}^{s} |f(x+rt)|\, dr$$

is the Hardy–Littlewood maximal operator in the direction t. The L^p $(1 \leqslant \infty)$ theory for M_Ω now follows immediately from the corresponding theory for the one-dimensional maximal function, and Minkowski's inequality for integrals:

$$\| M_\Omega(f)\|_p \leqslant \| M_\Omega^*(f)\|_p = \left\| \int_{S^{n-1}} \Omega(t)\, M^t(f)\, d\sigma(t) \right\|_p$$

$$\leqslant \int_{S^{n-1}} \Omega(t)\, \| M^t f \|_p\, d\sigma(t) \leqslant A_p \|f\|_p \int_{S^{n-1}} \Omega(t)\, d\sigma(t).$$

This answers our question in case $p > 1$. But for $p = 1$, the method of proof breaks down and the reason for this is that the space weak L^1 is not normed. (Recall that weak L^1 is the collection of all functions f on R_T satisfying $m\{x \mid |f(x)| > \alpha\} \leqslant A/\alpha$ for some A independent of α, and the smallest such A is referred to as the weak L^1 "norm" of f, $\|f\|_{WL^1}$, even though $\| \quad \|_{WL^1}$ is not a norm.)

In the case $p > 1$, the estimate

$$M_\Omega(f) \leqslant \int_{S^{n-1}} \Omega(t)\, M^t(f)\, d\sigma(t)$$

dominates $M_\Omega(f)$ by an average of functions with bounded L^p norms. But if we only know that $f \in L^1$, then $M^t f$ will have bounded weak L^1 norms, and it is not true that an average of functions each of which has weak L^1 norm $\leqslant 1$ is itself weak L^1. Nevertheless, we shall show, using a modification of the method of rotations, that if Ω satisfies an L^1 Dini condition, i.e., if

$$\| \Omega \|_{D^1} = \iint_{x,y \in S^{1-n}} \frac{|\Omega(x) - \Omega(y)|}{|x-y|^{n-1}}\, d\sigma(x)\, d\sigma(y) < \infty$$

then M_Ω is of weak type $(1, 1)$ with a weak-type bound no larger than $C = \|\Omega\|_{D^1}$ (C depending only on the dimension n).

Now, note that in the case $p > 1$, we relied upon the boundedness of M_*^Ω on L^p in order to establish the boundedness of M_Ω there. Before we investigate the L^1 behavior of M_Ω we shall show that the operator Mn can no longer be used to show the boundedness of M_Ω from L^1 to weak L^1. More precisely what will now be shown is that even if $\Omega(t) \equiv 1$ on S^{n-1}, the operator M_Ω^* is not weak type (1.1).[1]

[1] I wish to thank my friend Jean Clerc for pointing out this problem to me.

To show that the operator M^* which, at each point of R^2, averages the Hardy–Littlewood maximal operator in all the different directions is not weak type we make two observations. Note first that, as stated above, the reason that we cannot immediately conclude that M^* is weak type (1, 1) is that weak L^1 is not normed. And the nicest counterexample showing that weak L^1 is not normed is gotten by taking N copies of the kernel $1/|x|$ and spacing them evenly throughout the interval

$$[0, 1] : \left\| \sum_{k=1}^{N} \frac{1}{|x - k/N|} \right\|_{\mathrm{W}L^1} \sim N(\log N),$$

whereas

$$\sum_{k=1}^{N} \left\| \frac{1}{(x - k/N|} \right\|_{\mathrm{W}L^1} = N.$$

The second thing to keep in mind is that in order to get a function for which M^*f is large on a large set, the mass of f must be spread out in such a way that for x in some large set, there are many directions in which f looks large (this is not the case, for example, in the counterexample associated with the Kakeya problem where for each point in a large set, there is only one bad direction).

Taking these two observations as motivation, we consider a system of N^2 disks $D_{k,j}$ centered at the integer lattice points of the plane (j, k), $1 \leqslant j \leqslant N$, $1 < k \leqslant N$. Spread a unit mass uniformly on each of these disks, and draw all of the lines which pass through any fixed pair of the disks. Let C be the set of all points $(x, y) \in R^2$ with $0 \leqslant x, y \leqslant N$ and such that (x, y) lies on none of the lines we have constructed. Then if $z \in C$, then any point of one of the disks lies on a ray emanating from z and passing through no other disk. It follows that for $z \in C$,

$$M^* \left(\frac{1}{|D_{kj}|} \sum \chi_{D_{kj}} \right)(z) = \frac{1}{|D_{kj}|} \sum M^*(\chi_{D_{kj}})(z);$$

but

$$\sum \frac{1}{|D_{kj}|} M(\chi_{D_{kj}})(z) \geqslant A \log N.$$

If the size of the D_{kj} is sufficiently small, then $|C| \sim N^2$ and so $\| M^*(f)\|_{\mathrm{W}L^1}/\|f\|_1 \geqslant A \log N$ for $f = (1/|D_{kj}|) \sum \chi_{D_{kj}}$, proving that M^* is not weak type (1, 1).

As we have already stated, however, all is not lost. What is true is that for Ω just a bit better than L^1, M_Ω is an operator of weak type on L^1. To start our program off, assume that $\Omega = (\chi_{\cup I_k}) \mid \bigcup I_k \mid^{-1}$ for disjoint intervals I_k. (For convenience, we shall work in R^2, so that S^1 is identified with $[0, 2\pi)$.) Let us denote by T_k the wedge $\{z \in R^2 \mid z = re^{i\theta}, \theta \in I_k, 0 \leqslant r \leqslant 1\}$. Then we get the estimate

$$\frac{1}{r^2}\int_{|t|\leqslant r}\Omega(t)\,f(x+t)\,dt$$

$$=\frac{1}{|\bigcup I_j|}\sum\frac{1}{r^2}\int_{rT_k}f(x+t)\,dt-\sum\left(\frac{|T_k|}{|\bigcup I_j|}\right)\frac{1}{|rT_k|}\int_{rT_k}f(x+t)\,dt$$

$$\leqslant\sum\left(\frac{|T_k|}{|\bigcup I_j|}\right)M_{T_k}(f)(x),$$

where M_{T_k} is the maximal operator corresponding to averaging over dilates of T_k. The crucial fact which we use now is that the operators M_{T_k} are weak type (1, 1) with a weak-type bound independent of k. This is because maximal operators corresponding to the dilates of a convex body symmetric about 0 are uniformly weak type, and in the case of the T_k, the smallest such convex body C_k has measure comparable to T_k itself, so that $M_{T_k}\leqslant AM_{Ck}$ has measure comparable to T_k itself, so that $M_{T_k}\leqslant AM_{C_k}$.

Assume now that $\|f\|_1\leqslant 1$. Then $\|M_{T_k}(f)\|_{\mathrm{W}L^1}\leqslant C$ and we have the estimate $M_\Omega(f)\leqslant\sum(|I_k|/|\bigcup I_j|)\,M_{T_k}(f)$. If weak L^1 were normed we should have $\|M_\Omega(f)\|_{\mathrm{W}L^1}\leqslant C\sum(|I_k|/|\bigcup I_j|)=C$. This is not so, but instead, we may use a theorem of Stein and N.J. Weiss [10] which describes the way weak L^1 functions add:

THEOREM OF STEIN AND N. WEISS. *Let $f_1, f_2,\dots f_k,\dots$ be weak L^1 functions of norm $\leqslant 1$, and suppose that $0<a_k<1$ are such that $\sum a_k=1$ and $\sum a_k\log(1/a_k)=a<\infty$. Then the function $f=\sum a_kf_k$ satisfies $\|f\|_{\mathrm{W}L^1}\leqslant K(a+1)$ for some absolute constant K.*

The proof of this result can be found in [10]. Applying it to the preceding estimate, we arrive at $\|M_\Omega f\|_{\mathrm{W}L^1}\leqslant K\sum|I_k|\log(1/|I_k|)$. There is something unnatural about this estimate. It might well be that there are intervals J_k with $\bigcup J_k\supset\bigcup I_k$ and so that $\sum|J_k|\log(1/|J_k|)\leqslant\sum|I_k|\log(1/|I_k|)$. In fact, rather than our initial estimate, we can say that

$$\|M_{\chi_{\cup I_k}}f\|_{\mathrm{W}L^1}\leqslant K\inf_{\substack{\cup J_k\supset\cup I_k\\ J_k\,\text{intervals}}}\sum|J_k|\log\frac{1}{|J_k|}.$$

So, if for any subset, S, of $[0,2\pi]$ we let the entropy of S be the number

$$E(S)=\inf_{\substack{\cup J_k\supset S\\ J_k\,\text{intervals}}}\sum|J_k|\log\frac{1}{|J_k|},$$

then it follows that $\|M_{\chi_{S/|S|}}(f)\|_{\mathrm{W}L^1}\leqslant KE(S)\cdot\|f\|_1$. Now in order to generalize this result to arbitrary kernels $\Omega\geqslant 0$ such that $\int_{S^1}\Omega=1$, we introduce the notion of entropy of a function which parallels the concept of the L^1 norm of a function in the setting of Lebesgue measure. For $f \geqslant 0$ defined on $[0, 2\pi]$ let

$$\mathbf{J}(f) = \int_0^\infty E\{x \mid f(x) > \alpha\}\, d\alpha.$$

Then for $\Omega \geqslant 0$, with integral 1 over S^1, we have

$$\| M_\Omega(f) \|_{\mathrm{W}L^1} \geqslant K \mathbf{J}(\Omega) \, \| f \|_1 .$$

But before showing this, we wish to dispose briefly of a trivial matter. Technically speaking, in the proof we are about to give, and in any discussion of entropy in general, we should be sure to have $E(S) \geqslant c \mid S \mid$. If we let

$$E(S) = \inf_{\substack{S \subset \bigcup J_k \\ J_k \text{ intervals in } [0,2\pi]}} \sum \mid J_k \mid \log \frac{1}{\mid J_k \mid}$$

then $E(S)$ would be negative for every S, since $\log(1/2\pi) < 0$. We avoid this by defining a dilation, δ, of $[0, 2\pi]$ onto $[0, 1/e]$ and then for any $S \subset [0, 2\pi]$ we apply the definition of entropy given above to $\delta(S) \subset [0, 1/e]$ calling this quantity $E(S)$. In this way, entropy is monotone and positive, and we do in fact have $E(S) \geqslant c \mid S \mid$. During any subsequent discussion however, we choose to ignore this difficulty.

Now returning to the estimate $\| M_\Omega(f)_{\mathrm{W}L^1} \leqslant K \mathbf{J}(\Omega) \, \| f \|_1$, we select, for each $n \geqslant 1$ intervals $\{I_k{}^n\}_{k=1,2,\ldots}$ satisfying $\bigcup I_k{}^n \supset \{x \mid 2^n \leqslant \Omega(x) < 2^{n+1}\}$ and

$$\sum \mid I_k{}^n \mid \log \frac{1}{\mid I_k{}^n \mid} \leqslant E\{x \mid 2^n \leqslant \Omega(x) < 2^{n+1}\} + \left(\frac{1}{2^n}\right)^2.$$

We estimate as before:

$$\begin{aligned} &\frac{1}{r^2} \int_{|t| \leqslant r} \Omega(t) \, f(x+t) \, dt \\ &\qquad \leqslant \frac{1}{r^2} \int_{|t| \leqslant r} \sum_{n \geqslant 1} \sum_k 2^n \chi_{rT_k{}^n}(t) \, f(x+t) \, dt + \frac{1}{r^2} \int_{|t| \leqslant r} f(x+t) \, dt \end{aligned}$$

(here, as above, $T_k{}^n$ denotes the sector of the unit disk corresponding to the interval $I_k{}^n$)

$$\leqslant \sum_n 2^n \sum \mid I_k{}^n \mid M_{T_k{}^n}(f)(x) + M(f)(x),$$

where $M_{T_k{}^n}$ refers to the maximal operator which takes sup's of averages over dilates of $T_k{}^n$, and M is the classical two-dimensional Hardy–Littlewood maximal operator in R^2. Again using the theorem of Stein and N. Weiss, we have

$$\| M_\Omega(f)\|_{WL^1} \leqslant \Big(\sum_{n\geqslant 1} 2^n \sum_k | I_k^n | \log \frac{1}{| I_k^n |} + 1\Big) \cdot \| f \|_1$$
$$\leqslant \| f \|_1 \Big(\sum_{n\geqslant 1} 2^n E\{\Omega > 2^n\} + 2\Big)$$
$$\leqslant \| f \|_1 K \int_0^\infty E\{\Omega > \alpha\}\, d\alpha = KJ(\Omega) \| f \|_1 .$$

This should convince the reader that the entropy of the kernel Ω controls the operator M_Ω. To apply this result to singular integrals in a classical setting it is desirable to find a relationship between the entropy of Ω and some other quantity which arises in the theory of classical Fourier analysis. Since the entropy of a function is a quantity which depends both on size of functions and on the geometry of the sets where they are large, it is reasonable to compare entropy of a function with the sum of two norms, one measuring only size, the other measuring only smoothness. The expression we seek is the L^1 Dini norm, defined for a function f, on the circle by

$$\| f \|_{D^1} = \| f \|_{L^1} + \iint_{x,y\in[0,2\pi]} \frac{| f(x) - f(y)|}{| x - y |}\, dx\, dy.$$

What we intend to show is that functions having finite L^1 Dini norm (written $f \in D^1$) have finite entropy, and furthermore $J(f) \leqslant C \| f \|_{D^1}$. (Here we are only claiming that there is t $= f$ a.e. so that this inequality is satisfied.)

First, make the additional assumption that for some $S \subset [0, 2\pi]$, $f = \chi_S$. Then we must show that

$$E(S) \leqslant K \iint \frac{| \chi_S(x) - \chi_S(y)|}{| x - y |}\, dx\, dy$$

(also we may assume that $| S |$ is small; otherwise there is nothing to prove). Eliminating those points of S which are not points of density of S, if necessary, we may assume that all points of S are points of density. Now if $x \in S$ we may choose an interval I_x centered at x so that $| I_x \cap S | / | S | \geqslant \frac{1}{2}$ and so that if $\tilde{I}_x$ is any larger interval centered at x, then $| \tilde{I}_x \cap S | / | \tilde{I}_x | < \frac{1}{2}$. The I_x intervals form a cover of S, and we may choose a disjoint sequence of intervals $\{I_k\}$ with the property that $\bigcup_k 10I_k \supset S$ ($10I_k$ is the interval whose center is the same as I_k but whose length is 10 times that of I_k). Now, since

$$E(S) \leqslant \sum | 10I_k | \log \frac{1}{| 10I_k |} \leqslant 10 \sum | I_k | \log \frac{1}{| I_k |},$$

it suffices to prove that

$$| I_k | \log \frac{1}{| I_k |} \leqslant \int_{x \in I_k} \int_{y \in [0, 2\pi]} \frac{| \chi_S(x) - \chi_S(y) |}{| x - y |} \, dx \, dy, \qquad k \geqslant 1.$$

To prove this, take x to be the center point of I_k, and write

$$\begin{aligned}
&\int_{y \in [0, 2\pi], y \notin I_k} \frac{| \chi_S(x) - \chi_S(y) |}{| x - y |} \, dy \\
&\quad = \int_{y \in S \cap I_k} \frac{1}{| x - y |} \, dy = \int \frac{1}{| x - y |} \chi_{{}^cS \cap {}^cI_k}(y) \, dy. \\
&\quad = \int_{|I_k|/2}^{\pi} \frac{1}{r} \, dm\{y \in {}^cS \mid | x - y | \leqslant r\} \\
&\quad = \frac{| {}^cS |}{\pi} - \frac{| {}^cS \cap I_k |}{| I_k |} + \int_{|I_k|/2}^{\pi} m\{y \in {}^cS \mid | x - y | \leqslant r\} \frac{dr}{r^2} \\
&\quad \geqslant -1 + \int_{|I_k|/2}^{\circ} r \frac{dr}{r^2} \sim \log \frac{1}{| I_k |}.
\end{aligned}$$

that I_k is an interval centered at x and is maximal with respect to the property that $| {}^cS \cap I_k | / | I_k | \leqslant \frac{1}{2}$.)

Now if $c \in S \cap I_k$, but not necessarily at the center of I_k, we have for $y \in {}^cS \cap {}^cI_k$,

$$\frac{1}{| x - y |} \geqslant \frac{1}{2} \frac{1}{| x_c - y |},$$

where x_c is the center point of I_k. It follows that

$$\int_{y \in {}^cS \cap {}^cI_k} \frac{| \chi_S(x) - \chi_S(y) |}{| x - y |} \, dx \, dy \geqslant C \log \frac{1}{| I_k |},$$

and because at least one-half of I_k is taken up by points $x \in S$,

$$\int_{x \in S \cap I_k} \int_{y \in {}^cI_k \cap {}^cS} \frac{| \chi_S(x) - \chi_S(y) |}{| x - y |} \, dx \, dy \geqslant C \, | I_k | \log \frac{1}{| I_k |}.$$

Now, because the I_k are disjoint,

$$\begin{aligned}
\int_{x, y \in [0, 2\pi]} \int \frac{| \chi_S(x) - \chi_S(y) |}{| x - y |} &\geqslant \sum_k \int_{I_k} \int_{y \in [0, 2\pi]} \frac{| \chi_S(x) - \chi_S(y) |}{| x - y |} \, dx \, dy \\
&\geqslant C \sum | I_k | \log \frac{1}{| I_k |}
\end{aligned}$$

and we have shown that, for $f = \chi_S$, $\mathbf{J}(f) \leqslant C \, \| f \|_{D^1}$.

Our proof of this last fact extends immediately to a somewhat more general situation. Suppose $f(x) = \sum_{k=1}^{N} 2^k \chi_{S_k}(x)$; because for $x \in S_k$, $y \in S_j$, $j \neq k$, $|f(x) - f(y)| \geqslant f(x)$, our argument above gives $\mathrm{J}(f) \leqslant C \| f \|_{D^1}$. Since simple functions, i.e., functions of the form $f(x) \sum \alpha_k \chi_{S_k}(x)$ (we assume here that $f \geqslant 0$, $\int_0^{2\pi} f = 1$), are already general enough for our purposes, what we shall now do is prove that to every simple f, there corresponds a simple f^* taking only finitely many of the values 2^k, $k \geqslant 0$, and satisfying $\| f^* \|_{D^1} \leqslant C \| f \|_{D^1}$ with $f(x)/2 \leqslant f^*(x) \leqslant 2f(x)$ for every x where $f(x) > 1$. Taking this for granted, we have $\mathrm{J}(f) \leqslant 2\mathrm{J}(f^*) \leqslant 2C_1 \| f^* \|_{D^1} \leqslant 2C_1C_2 \| f \|_{D^1}$, which is what we want.

The procedure we follow in order to construct a dyadic f^*, given a simple function $f = \sum \alpha_i \chi_{S_i}$ $(\alpha_i \geqslant 0)$, is quite easy. We begin changing f by redefining f to be equal to 1 where it was originally $\leqslant 1$. Call this new function f_1. Then, clearly, $\| f_1 \|_{D^1} \leqslant \| f \|_{D^1}$. Now we change f_1 slightly to produce a new function f_2 as follows: Let α_{k_1} be the smallest value of $f > 1$. Assume with no loss in generality, that $1 < \alpha_{k_1} < 2$. Then on the set S_{k_2} where $f_1 = \alpha_{k_1}$ we shall set $f_2 = \alpha_{k_2}$ or 1 (here α_{k_2} is the next smallest value assumed by f_1) depending on whether

$$\sum_{\alpha_{k_j} > \alpha_{j_1}} \int_{x \in S_{k_1}} \int_{y \in S_{k_j}} \frac{1}{|x-y|}\, dx\, dy \qquad \text{or} \qquad \sum_{\alpha_{k_j} < \alpha_{k_1}} \int_{x \in S_{k_1}} \int_{y \in S_{k_j}} \frac{1}{|x-y|}\, dx\, dy$$

is larger, respectively. (Here, $f = \alpha_{k_j}$ on S_{k_j} is the notation we are using.) And outside of S_{k_1} set $f_2 = f_1$. Then we have $\| f_2 \|_{D^1} \leqslant \| f \|_{D^1}$. Now, to get f_3 we alter f_2 as follows. Suppose α_{k_3} is the third smallest value <1 which is assumed by f. Again, for convenience, assume $\alpha_{k_3} < 2$. Where $f_2 \neq \alpha_{k_2}$ we let $f_3 = f_2$. On the set where $f_2 = \alpha_{k_2}$ we define f_3 to be either α_{k_3} of 1 depending on whether

$$\sum_{\alpha_{k_j} > \alpha_{k_2}} \int_{x \in S_{k_2}} \int_{y \in S_{k_j}} \frac{1}{|x-y|}\, dx\, dy \qquad \text{or} \qquad \sum_{\alpha_{k_j} < \alpha_{k_2}} \int_{x \in S_{k_2}} \int_{y \in S_{k_j}} \frac{1}{|x-y|}\, dx\, dy$$

is larger. (If $\alpha_{k_3} \geqslant 2$, we would move f_2 up to 2 or down to 1 depending on which of the above integrals is larger.) Then $\| f_3 \|_{D^1} \leqslant \| f_2 \|_{D^1}$, and continuing on in this way we have the required dyadic f^* after finitely many steps.

Let us review our position. We now know that the operator M_Ω is weak type (1, 1) if the kernel Ω satisfies an L^1 Dini condition. The next step in our program is to apply this result to the theory of singular integrals in order to give a new proof of the M. Weiss–Calderón–Zygmund theorem, which reads as follows:

Theorem of M. Weiss, Calderón, and Zygmund. *Suppose that Ω is a function on R^n, homogeneous of degree 0, satisfying an L^1 Dini condition when restricted to S^{n-1} and having mean value 0 on the sphere.*

Then the singular integral operator

$$T(f)(x) = \lim_{\epsilon\to 0} \int_{|y|>\epsilon} \frac{\Omega(y)}{|y|^n} f(x-y)\, dy$$

(*this limit exists a.e. for* $f \in L^1(R^n)$) *is of weak type* (1, 1).

This theorem was first proved in [12]. Again, for convenience, alone, we work in R^2, so that the sphere has a convenient parameterization. To prove this theorem, let us prove a lemma.

LEMMA. *If f is a positive function on $T^1 = [0, 2\pi)$ which is in the Dini class, then* $f \in L$ ($\log^+ L$) and

$$\int_{T^1} |f(\theta)| \log^+ |f(\theta)|\, d\theta \leqslant C \|f\|_{D^1}.$$

There is a complex-variables proof of this lemma in [12]. The authors then proceed to establish the general lemma in R^n by induction. The advantage of our entropy-theoretic proof for $n = 1$ given here is that it is a real variable proof and generalizes immediately to R^n, $n > 1$.

Proof of the lemma. We know that if $\|f\|_{D^1} < \infty$ then $J(f) < \infty$ and $J(f) \leqslant C\|f\|_{D^1}$. We observe now that if $J(f) < \infty$ then $\int |f| \log^+ |f| \leqslant C' \|f\|_{D^1}$. The proof that $\int |f| \log^+ |f| \leqslant CJ(f)$ is a simple consequence of the fundamental property of entropy, namely, that for a fixed Lebesgue measure, the entropy of a set is minimized when the set is an interval. In other words $E(S) \geqslant |S| \log(1/|S|)$. Then

$$J(f) = \int_0^\infty E\{x \mid f(x) > \alpha\}\, d\alpha \geqslant C \sum 2^k E\{f > 2^k\}$$

$$\geqslant \sum_k 2^k |\{x \mid f(x) > 2^k\}| \log \frac{1}{|\{x \mid f(x) > 2^k\}|}.$$

If we assume that

$$\int f = 1, |\{f > 2^k\}| \leqslant 1/2^k, \text{ so } \log(1/|f > 2^k|) \geqslant \log 2^k$$

and

$$J(f) \geqslant C \sum (2^k \log 2^k) |f > 2^k| \geqslant C' \int |f| \log^+ |f|.$$

With this lemma out of the way, we are now ready to prove the theorem of M. Weiss, Calderón, and Zygmund.

Let $f \in L^1$ and let $\alpha > 0$. Apply the Calderón-Zygmund decomposition (see [2]) to $|f|$ at height α. So $R^n = F \cup G$, where $|f| \leqslant \alpha$ a.e. on F and $G = \bigcup Q_k$,

Q_k disjoint cubes with $(1/|Q_k|)\int_{Q_k}|f(x)|\,dx \sim \alpha$. As usual, define the good and bad functions as follows: $f = g + b$: $g(x) = f(x)$ on F, and $g(x) = (1/|Q_k|) \int_Q f(y)\,dy$ on Q_k. If we let

$$T_\epsilon(f)(x) = \int_{|y|>\epsilon} \frac{\Omega(y)}{|y|^n} f(x-y)\,dy \qquad \text{and} \qquad T^*(f)(x) = \sup_{\epsilon>0} |T_\epsilon(f)(x)|,$$

then T^* is bounded on L^2; this result is valid whenever the kernel Ω has mean value 0 on the sphere and is $L\ \log^+ L)(S^{n-1})$. For this result we refer to the work of Calderón and Zygmund [3]. By the lemma, Ω is $L(\log L)$ since it is in the Dini class. So

$$m\{x \mid T^*g(x) > \alpha\} \leqslant \frac{\|T^*g\|_2^2}{\alpha^2} \leqslant A^2 \frac{\|g\|_2^2}{\alpha^2} \leqslant CA^2 \frac{\|f\|_1}{\alpha}.$$

Now, in order to handle $\{T^*b > \alpha\}$, we shall need to use the fact, proved in [11] that if $K(x) = \Omega(x)/|x|^n$ then

$$\int_{|x|>2|y|} |K(x+y) - K(x)|\,dx \leqslant C,$$

for each $y \in R^n$, whenever $\|\Omega\|_{D^1} < \infty$. (Here

$$\|\Omega\|_{D^1} = \int_{S^{n-1}} |\Omega(x)|\,d\sigma(x) + \int_{x,y\in S^{n-1}}\int \frac{|\Omega(x) - \Omega(y)|\,d\sigma(x)\,d\sigma(y)}{|x-y|^{n-1}}.$$

Let $\tilde{Q}_k$ be a cube with the same center as Q_k, only with sides 10 times as long. Then for a fixed $\epsilon > 0$ and $x \notin \bigcup \tilde{Q}_k$, consider two classes of cubes: $\mathcal{O}\!\mathcal{l}_1$ is the family of cubes Q_k intersecting the sphere $\{y \mid |x-y| = \epsilon\}$ and $\mathcal{O}\!\mathcal{l}_2$ is the family of all other Q_k. Then

$$T_\epsilon(b)(x) = \sum_{Q_k\in\mathcal{O}\!\mathcal{l}_1} \int_{Q_k\cap{}^cB(x;\epsilon)} K(x-y)\,b(y)\,dy + \sum_{Q_k\in\mathcal{O}\!\mathcal{l}_2} \int_{Q_k} K(x-y)\,b(y)\,dy$$

(here $B(x;\epsilon)$ is the ball in R^n centered at x of radius $\epsilon > 0$)

$$= \sum_{Q_k\in\mathcal{O}\!\mathcal{l}_1} \int_{Q_k\cap{}^cB(x;\epsilon)} K(x-y)\,b(y)\,dy$$

$$+ \sum_{Q_k\in\mathcal{O}\!\mathcal{l}_2} \int_{Q_k} [K(x-y) - K(x-y_k)]\,b(y)\,dy$$

(here y_k is the center of Q_k).

So we see that

$$|T_\epsilon b(x)| \leqslant \sum_{Q_k\in\mathcal{O}\!\mathcal{l}_1} \int_{Q_k\cap{}^cB(x,\epsilon)} \frac{|\Omega(x-y)|}{|x-y|^L} |b(y)|\,dy$$

$$+ \sum_{\text{all } k} \int_{Q_k} |K(x-y) - K(x-y_k)|\,|b(y)|\,dy = \beta_\epsilon(x) + \gamma(x).$$

(β_ϵ is the first term of the preceding sum, and γ the second.) Now, as always, $\int_{c(\cup\tilde{Q}_k)} \gamma(x)\, dx \leqslant C \| f \|_1$. We need only handle β_ϵ, which is the critical term reflecting the truncation at ϵ. And the point is that each cube $Q_{1\,K}$ K_1 intersects $\{y \mid | x - y | = \epsilon\}$, but $x \in \tilde{Q}_k$, so that the diameter of Q_k is small compared to ϵ. Therefore $| x - y | \sim \epsilon$ for any $y \in Q_k$, for each $Q_k \in \mathcal{O}_1$. With this observation it is clear that

$$\beta_\epsilon(x) \leqslant \frac{C}{\epsilon^L} \int_{|x-y| \leqslant C\epsilon} | \Omega(x-y)| \, | b(y)| \, dy \leqslant C' M_\Omega(b)(x),$$

and M_Ω, of course, does not depend on ϵ. So

$$\begin{aligned} |\{T^*(f) > \alpha\}| &\leqslant |\{T^*(g) > \alpha/2\}| + |\{T^*(b) > \alpha/2\}| \\ &\leqslant C \| f \|_{1/\alpha} + |\{\gamma > \alpha/4\}| + |\{M_\Omega(b) > \alpha/4\}| \\ &\leqslant C' \| f \|_{1/\alpha} \end{aligned}$$

(because for Ω in the Dini class, M_Ω is weak type (1, 1) so that $| \{M_\Omega(b) > \alpha\}| \leqslant C \| b \|_{1/\alpha} \leqslant C' \| f \|_{1/\alpha}$). whis establishes that T^* is of weak type (1, 1) as desired, and finishes the proof.

2. The Set Function Entropy

What kind of set function is entropy? This is the major question discussed throughout this section. Later, we shall make use of the various properties of entropy as we continue to apply the concept to various problems in analysis.

First, let us repeat the definition. Let $S \subset Q_0$ in R^n where Q_0 is the unit cube. Then

$$E(S) = \inf_{\substack{S \subset \bigcup Q_k \\ Q_k \text{ subcubes of } Q_0}} \sum_{k \geqslant 1} | Q_k | \log \frac{1}{| Q_k |},$$

where $| Q |$ refers to the n-dimensional Lebesgue measure in R^n. It is clear right away that entropy has the following simple properties:

(1) $E(S) \geqslant 0$, $S \subset Q_0$,

(2) $E(S) \leqslant E(T)$, $S \subset T \subset Q_0$,

(3) $E(\bigcup_k S_k) \leqslant \sum_k E(S_k)$.

There is a fourth property of entropy which is intimately connected with the idea behind this concept:

(4) If $S \subset Q_0$, and $Q \subset Q_0$ are sets with the same Lebesgue measure and

Q is a cube, then $E(S) \geqslant E(Q)$. This is true because if Q_k are cubes covering S then we may assume each Q_{k_0} is smaller than Q (otherwise

$$\sum |Q_k| \log \frac{1}{|Q_k|} \geqslant |Q_{k_0}| \log \frac{1}{|Q_{k_0}|} \geqslant |Q| \log \frac{1}{|Q|},$$

and we are finished). But then

$$\sum |Q_k| \log \frac{1}{|Q_k|} \geqslant \left(\sum |Q_k|\right) \log \frac{1}{|Q|} \geqslant |Q| \log \frac{1}{|Q|}.$$

A property of entropy which is useful in situations of monotone convergence is

(5) If S_k is a sequence of subsets of Q_0 with $S_k \subset S_{k+1}$, then $E(\bigcup_k S_k) = \lim_{k\to\infty} E(S_k)$. For the proof, we refer the reader to [6, Chap. II].

At this point, we shall consider a basic example which will be of some use later. Let $\delta > 0$ be small, and N a large positive integer, with $N\delta < 1$. Set

$$S_{N,\delta} = \bigcup_{k=0,1,\ldots,[1/N\delta]} [kN\delta, kN\delta + \delta].$$

Then the entropy of $S_{N\delta}$ is given by $E(S_{N\delta}) \sim \min(1, (1/N) \log(1/\delta)$ (as usual $A \sim B$ means that thrre are constants $0 < c$, $C < \infty$ so that $cB \leqslant A \leqslant CB$). Before provising this, notice what this really says. In order to find the entropy of a set we must find this minimizing cover for $S_{N\delta}$ we need only consider two covers: the cover consisting of the interval $[0, 1]$, and the cover consisting of the component intervals of $S_{N\delta}$. The one of these two covers with the smaller sum $\sum |I_k| \log(1/|I_k|)$ is the cover which essentially minimizes this sum as we vary over *all* possible covers.

To prove this, we first establish the notation that each $S_{N\delta}$ is made up of $[1/N\delta]$ component intervals $I_k^{N\delta}$ or for simplicity I_k. Let J_k cover $S_{N\delta}$. Then there are two possibilities. First suppose that at least half of the I_k have the property that they are contained in a J_k which does not intersect any other I_k. In this case we have

$$\sum |J_k| \log \frac{1}{|J_k|} \geqslant \frac{1}{2} \left[\frac{1}{N\delta}\right] \delta \log \frac{1}{\delta} \sim \frac{1}{N} \log \frac{1}{\delta}.$$

In the second case, we have at least half of the I_k contained in coverging intervals J_k each of which covers at least 2 different I_k. Now the key point is that the ratio of the Lebesgue measure of the smallest interval containing at least m consecutive I_k to the measure of those m I_k is independent of m as long as $m \geqslant 2$. Therefore, since the J_k which contain at least two distinct I_k cover at least half the I_k, the Lebesgue measure of the union of these J_k must be at least $\frac{1}{2}$.

In either case

$$\sum |J_k| \log \frac{1}{|J_k|} \geqslant \min \left(\frac{1}{N} \log \frac{1}{\delta}, 1\right).$$

This proves that up to a constant factor

$$E(S_{N\delta}) \geqslant \min \left(\frac{1}{N} \log \frac{1}{\delta}, 1\right),$$

and the reverse inequality is obvious.

The example we have just considered carries with it two important messages. First, entropy is not at all comparable with Lebesgue measure, and second entropy is not invariant under dilations. In fact if $\delta = e^{-N}$, we see that although $S_{N\delta}$ has Lebesgue measure $1/N$, it has entropy ~ 1. From here it is clear that we can construct a function with infinite entropy whose Lebesgue distribution function, though never 0, approaches 0 as rapidly as we like. The idea of having a function near L^∞ with infinite entropy is reasonable, since entropy involves more than just the size of a function; it involves its geometry as well. As for the failure of entropy to be dilation invariant, this is shown by the fact that $\chi_{S_{N\delta}}$ is a dilate mod 1 of the function $\chi_{[0,1/N]}$ which has small entropy though $\chi_{S_{N\delta}}$ need not. This will be important later on when we consider the action of operators which commute with dilations, on entropy which does not commute.

Now, in considering entropy, it is a good idea to keep in mind some of the rather similar properties of capacity. So let us introduce a kernel $K(x)$ defined on $[-\pi, \pi]$ which is positive, even, and such that $\int_{-\pi}^{\pi} K(x)\, dx < \sim$. Then the capacity $C_K(S)$ of a set S with respect to the kernel K is defined as follows:

$C_K(S) = \sup \mu([-\pi, \pi])$ where the sup is taken over all positive measures μ such that μ is supported inside S and so that $K * \mu(x) \leqslant 1$ a.e. on S. Then C_K is a positive monotone function of S which, like entropy is subadditive. Many of the results and estimates which hold for E have analogous (but inequivalent) formulations in terms of capacity. The major difference seems to be that somehow, because capacity is defined in terms of convolutions, it is more intimately connected with the Fourier analysis of L^2 spaces than is entropy. On the other hand, sharp L^1 estimates seem to call for entropy rather than capacity. But what exactly is the relationship between these two set functions? To answer this question, we should keep in mind that for every positive, increasing, concave function on [0, 1], say $\varphi(x)$, we could have defined a notion of entropy with respect to φ:

$$E_\varphi(S) \inf_{\substack{S \subset \bigcup I_k \\ I_k \text{ intervals}}} \sum \varphi(|I_k|).$$

And the point is that to every K there corresponds in a natural way a φ so that E_φ and C_k are closely related.

The correspondence is given by

$$\varphi(x) = x \left[\int_0^x K(t)\, dt\right]^{-1}.$$

In particular, the kernel corresponding to the entropy originally introduced is

$$K(x) = 1/|x| \left(\log \frac{1}{|x|}\right)^2,$$

and with respect to this kernel intervals I_k have capacity $|I| \log(1/|I|)$. For any set S and a covering of S by intervals I_k with the property that $\sum |I_k| \log(1/|I_k|) < E(S) + \epsilon$ we have $C(S) \leqslant \sum C(I_k) = \sum |I_k| \log(1/|I_k|) < E(S) + \epsilon$ so that $C(S) \leqslant E(S)$. The reverse inequality is not valid. In fact there is a necessary and sufficient condition which one can give for the vanishing of capacity to imply the vanishing of entropy:

If $\int_0^1 K(x)\, d\varphi(x) < \infty$ then $C_K(S) = 0$ implies $E_\varphi(S) = 0$. But if $\int_0^1 K(x)\, d\varphi(x) = \infty$ there exist compact sets F for which $C_K(F) = 0$, but $E_\varphi(F) > 0$. (For these results see the work of Carleson [5], and Taylor [11].)

In our case, where

$$K(x) = 1/|x| \left(\log \frac{1}{|x|}\right)^2$$

and $\varphi(x) = x \log(1/x)$ the integrand $K(x)\, d\varphi(x) = (1/x)\, dx$, and the fact that $\int_0^1 (1/x)\, dx$ just barely diverges shows the closeness of these two set functions. As we explore the action of some familiar operators on entropy, we shall parallel this with a discussion of the corresponding known results about capacity.

Now, at this point, there are three topics which will be treated in the remainder of this section. Very briefly, the relationship between entropy and Hausdorff measure will be mentionad. Next, there will be some consideration of L^p space with respect to entropy. And finally, we shall take up the important question of whether entropy of functions is a normed quantity. Then we shall use this information in the next and last section to obtain some estimates of a classical nature.

Proceeding with the program, it should be pointed out that there is a close connection between E_φ and the corresponding Hausdorff measure H_φ defined as follows: Suppose $\delta > 0$. Define $H_\varphi{}^\delta(S)$ as the inf of all sums $\sum_k \varphi(|I_k|)$, where the inf is taken over all sequences of intervals I_k which cover S and such that $|I_k| < \delta$ for all k. Then $H_\varphi{}^\delta(S)$ increases as δ increases and therefore $H_\varphi(S) = \lim_{\delta\to 0} H_\varphi{}^\delta(S)$ exists and is called the Hausdorff measure of S relative to φ. Although H_φ and E_φ are completely different as set functions (H_φ is additive on Borel sets), still $H_\varphi(S)$ vanishes iff $E_\varphi(S) = 0$. See [5]. This reduces the study of properties which happen except on a set of 0 Hausdorff measure to a quantitative problem about entropy. So, as an example, in the next section we shall establish the estimate $E\{Mf > \alpha\} \leqslant (C/\alpha) \|f\|_{D^1}$ where f is in the L^1 Dini class on $[0, 2\pi]$, and M is the Hardy–Littlewood maximal operator. Just as the usual maximal theorem implies differentiability of the integral except on a set of Lebesgue measure 0, this estimate gives differentiation except on a set of 0 entropy, or what is the same thing, on a set of 0 Hausdorff measure.

Next, let us ask about the L^p spaces taken with respect to the set function E.

The natural definition to make here is that $L^p(E)$ is the class of functions f defined on $[0, 2\pi]$ satisfying

$$\left(p \int_0^\infty \alpha^{p-1} E\{|f| > \alpha\}\, d\alpha\right)^{1/p} = \mathrm{J}_p(f) < \infty.$$

Then $\mathrm{J}(|f|^p) = \int_0^\infty E\{|f| > \alpha^{1/p}\}\, d\alpha = p\int_0^\infty E\{|f| > \alpha\}\, \alpha^{p-1}\, d\alpha = \mathrm{J}_p{}^p(f)$. So, from $\mathrm{J}_p(f) = \mathrm{J}(|f|^p)^{1/p}$ we can proceed formally exactly as in the case of Lebesgue measure to prove Hölder's inequality and Minkowski's inequality provided that we can show that $L^1(E)$ is normed. We shall show in a while that in one dimension this is the case, and in R^n this is very nearly the case. In any event, everything goes along smoothly until we try to extend the inequality

$$\mathrm{J}(f) < C \|f\|_{D^1}$$

to L^p, $p > 1$.

If it were the case that $\mathrm{J}_2(f) \leqslant C \|f\|_{D^1}$, where

$$\|f\|_{D^2}^2 = \|f\|_{L^2}^2 + \iint_{x,y\in T} \frac{|f(x) - f(y)|^2}{|x - y|}\, dx\, dy,$$

then it would also have to be true that

$$E(S) \leqslant \mathcal{C}C_K(S),\ K(x) = 1\Big/|x| \left(\log \frac{1}{|x|}\right)^2.$$

In fact it is known (see [1]) that the following set function is equivalent to capacity (we say that two set functions $\alpha(S)$ and $\beta(S)$ are equivalent provided there are constants c_1 and c_2 so that $c_2\beta(S) \leqslant \alpha(S) \leqslant c_1\beta(S)$):

$$C_{k,2}(S) = \inf \|f\|_{L^2}^2,$$

where the inf is taken over all f so that $f * K \geqslant 1$ on S.

But then given $\epsilon > 0$ we may choose $F \geqslant 1$ on S with $\|F\|_{D^2} < C_{k,2}(S) + \epsilon$. On the other hand, since $F \geqslant 1$ on S, $\mathrm{J}_2(F) \geqslant E(S)$. If we assume $\mathrm{J}_2(F) \leqslant C\|F\|_{D^2}$, then $E(S) \leqslant C_{k,2}(S) \sim C_k(S)$, and this is not the case. Thus, the L^p $(p > 1)$ theory of entropy fails to work as well as the L^1 theory, and when we encounter L^p estimates we shall use capacity instead.

The last part of our program in this section involves investigating whether or not J is a norm. The claim is going to be that in one dimension entropy is normed, and while it is not a norm for functions of several variables it is equivalent to a norm.

In one dimension, suppose we take two functions f and $g \geqslant 0$. Let $\eta > 0$. Let $\epsilon < 0$ be small so that

$$\sum_{k=0}^{\infty} \epsilon E\{f > k\epsilon\} < \mathrm{J}(f) + \eta,$$

$$\sum_{k=0}^{\infty} \epsilon E\{g > k\epsilon\} < \mathrm{J}(g) + \eta.$$

Cover the set $\{f > k\epsilon\}$ by intervals $\{I_{k,m}\}_{m=1,2\ldots}$ and the set $\{g > k\epsilon\}$ with intervals $I_{k,m}$ so that $\sum_m |I_{km}| \log(1/|I_{km}|) < E\{f > k\epsilon\} + 2^{-(k+1)}\eta$ and with a similar inequality for the $J_{k,m}$.

Consider the series

$$\sum_{k,m} \epsilon\chi_{I_{km}} + \sum_{k,m} \epsilon\chi_{J_{km}}.$$

If we can show that

$$\begin{aligned} \mathrm{J}\Big(\sum_{k,m\leqslant N} \epsilon\chi_{I_{km}} + \sum_{k,m\leqslant N} \epsilon\chi_{J_{km}}\Big) &\leqslant \epsilon\Big(\sum_{k,m\leqslant N} E(I_{km}) + E(J_{km})\Big) \\ &\leqslant J(f) + J(g) + 4\eta, \end{aligned} \tag{*}$$

then taking the limit as $N \to \infty$ by the monotone convergence theorem for entropy, we have $J(f+g) \leqslant J(f) + J(g)$. So all that remains to be proved is inequality (*), which says that

$$\mathrm{J}\Big(\sum_{k=1}^{N} \chi_{I_k}\Big) \leqslant \sum_{k=1}^{N} \mathrm{J}(\chi_{I_k}).$$

To prove this, note that if we take an interval, I, and chop it up into two intervals of lengths $\theta\,|I|$ and $(1-\theta)|I|$ and sum the entropies of the two pieces, then a trivial computation shows that this sum increases as θ varies from 0 to $\frac{1}{2}$, reaches a maximum there, and then decreases from $\frac{1}{2}$ to 1. Consider now intervals I and J which overlap. Then if we take an interval K of length $|I| + |J|$ and decompose it into intervals of lengths $|I|$ and $|J|$ this can be viewed as a decomposition with $\theta \geqslant \frac{1}{2}$. Decomposing K into an interval of length $|I \cup J|$ and one of length $|I \cap J|$, this decomposition corresponds to taking a $\theta' > \theta \geqslant \frac{1}{2}$, and so

$$E(I \cup J) + E(I \cap J) \leqslant E(I) + E(J), \qquad I, J \text{ intervals.} \tag{\#}$$

Now let us introduce just a bit of terminology. Suppose that we have a family of intervals $\{I_k\}$. If I^1 and I^2 belong to this family, and $I^1 \cap I^2 \neq 0$ we shall define a new family of intervals called "a simple replacement" of $\{i_k\}$ by keeping all the I_k in our family except for I^1 and I^2, and replacing these by the intervals $I^1 \cap I^2$ and $I^1 \cup I^2$. Then we have the following: Given any finite sequence of intervals $\{I_1\}_{k=1,2,\ldots,N}$ by a finite number of simple replacements we can replace the sequence $\{I_k\}$ with a sequence $\{I_k'\}$ having the property that given any pair of primed intervals I_j' and I_k' either $I_j' \cap I_k' = \varnothing$, $I_j' \subset I_k'$ or $I_k' \subset I_j'$. With this taken for granted, it follows that

$$\mathrm{J}\left(\sum \chi_{I_k}\right) \leqslant \sum \mathrm{J}(\chi_{I_k}).$$

In fact, because $\{I_k''\}$, a simple replacement of $\{I_k\}$, implies $\sum \chi_{I_k''} = \sum \chi_{I_k}$, we have

$$\mathrm{J}\left(\sum \chi_{I_k}\right) = \mathrm{J}\left(\sum \chi_{I_k'}\right) = \sum_{n=1}^{N} E\left(\sum \chi_{I_k'} = n\right).$$

But now the I_k' have the property that $\{\sum \chi_{I_k'} = n\}$ is actually a disjoint union of certain of the I_k' so that

$$\mathrm{J}\left(\sum \chi_{I_k}\right) = \sum_{n=1}^{N} E\left(\sum \chi_{I_k'} = n\right) \leqslant \sum E(I_k') \leqslant \sum E(I_k) = \sum \mathrm{J}(\chi_{I_k}).$$

(The last inequality follows from the fact that in passing to a simple replacement, the quantity $\sum E(I_k)$ is reduced because of inequality (#).) So to prove that in one dimension entropy is normed, we need only show that by simple replacements we can pass from an arbitrary finite collection $\{I_k\}$ of intervals to a collection $\{I_k\}$ so that any two intervals from the $\{I_k'\}$ have either one contained in the other, or are disjoint.

We prove this by induction on the number of intervals, n, in the family $\{I_k\}$. If $n = 1$, there is nothing to prove. Assume that the claim is true for n intervals, and consider a collection of $n + 1$ intervals. Now, there are two possibilities. In the first case, there is a sequence $I_{k_1}, I_{k_2}, \ldots, I_{k_m}$ so that $I_{k_1} \cap I_{k_2} \neq \varnothing$, $I_{k_2} \cap I_{k_3} \neq \varnothing, \ldots, I_{k_{m-1}} \cap I_{k_m} \neq \varnothing$, and such that if $I_{k_1} = (\alpha_1, \beta_1)$ and $I_{k_m} = (\alpha_m, \beta_m)$ then $\bigcup_{k=1}^{n+1} I_k \subset (\alpha_1, \beta_m)$. In this case we apply a simple replacement to the pair (I_{k_1}, I_{k_2}), then to the pair $(I_{k_1} \cup I_{k_2}, I_{k_3})$, then to $(I_{k_1} \cup I_{k_2} \cup I_{k_3}, I_{k_4})$, etc. Finally, at the last step, we shall have among the new intervals the interval (α_1, β_m) and there will be only n intervals inside it. By induction we may apply the replacement process finitely many times to arrive at a family of intervals $\{I_k''\}$ having the desired property. But then the I_k along with the interval (α_1, β_m) also has this property and we are finished with the proof in the first case. In the second case, that is, in the event that we are not in the first case, then there are intervals $I_{k_1}, I_{k_2}, \ldots, I_{k_m}$ so that if $I_{k_j} = (\alpha_j, \beta_j)$ then $\alpha_j < \alpha_{j+1} < \beta_j <$

β_{j+j} and so that no interval among the I_k intersects (α_1, β_m) without being contained in it. Just as in the first case, a finite number of replacements gets us to where (α_1, β_m) is one of the intervals of our family and inside of (α_1, β_m) we have $\leqslant n$ intervals. And in addition, all other intervals are disjoint from (α_1, β_m). So, just as before, we may apply the induction hypothesis to the intervals contained in (α_1, β_m) and to the intervals disjoint from (α_1, β_m) to obtain a family of intervals with the necessary property. This finishes the proof in the second case.

Now that we know that $J(f+g) \leqslant J(f) + J(g)$ for functions defined on $[0, 1]$, the natural question to ask is whether this inequality extends to functions on $Q_0 \subset R^n$. This is not the case, and the reason is that it is not true that

$$E(Q_1 \cap Q_2) + E(Q_1 \cup Q_2) \leqslant E(Q_1) + E(Q_2)$$

if Q_1 and Q_2 are n-dimensional cubes.

If we think about the proof in one dimension, we see that the basic idea was to prove that

$$J\left(\sum \chi_{I_k}\right) \leqslant \sum J(\chi_{I_k}) \qquad \text{for intervals } I_k \,. \tag{!}$$

And the inequality was proved by reducing an arbitrary finite collection of intervals to a collection having the special property that no two intervals in the collection intersect unless one is contained in the other. For such a collection the inequality (!) is quite simple. So, in n-dimensions, where the replacement process breaks down, the thing to do is to restrict our attention to a class of cubes already having this special property—the dyadic cubes. So if, in R^n, inside the unit cube, we define entropy by taking

$$E(S) = \inf_{\substack{S \subset \bigcup_k Q_k \\ Q_k \text{ dyadic cubes}}} \sum_k |Q_k| \log \frac{1}{|Q_k|}$$

then this new set function is equivalent to the old one and with respect to this definition, entropy of functions is normed.

In this section our attention will focus on applying what we know about entropy to problems is classical analysis.

3. Applications of Entropy

First we shall consider a problem concerning the nonincreasing rearrangement of a function. Although we restrict our attention to the one variable case, our methods extend easily to R^n, where we must then consider the radial rearrangement of f which decreases as the distance from the origin increases.

It is well known that given f, a function on $[0, 1]$, there is a nonincreasing function f^* on $[0, 1]$ which has the same distribution function as f. One natural question which can be asked about f^* is "What is the relationship between the smoothness of f and that of f^*?" Garsia [7] has studied this question using ingenious combinatorial arguments. Here we shall use entropy to show that

$$\|f^*\|_{D^1} \leqslant C \|f\|_{D^1}.$$

To prove this, we require the following lemma:

LEMMA. *For a decreasing function, f, the Dini norm is equivalent to the entropy.*

Proof. Take a function, f, decreasing on $[0, 1]$. Just as in the proof that $J(f) \leqslant C\|f\|_{D^1}$, we may find a function g which assumes only the values $0, 1, 2, 4, \ldots, 2^k, \ldots$, such that g is decreasing, $f/2 \leqslant g < 2f$, where $f \geqslant 1$, and this time, such that $\|f\|_{D^1} \leqslant C\|g\|_{D^1}$. If we show that $=g\|_{D^1} \leqslant CJ(g)$ then $\|f\|_{D^1} \leqslant C\|g\|_{D^1} \leqslant C'J(g) \leqslant C''J(f)$. And combining this with the inequality $J(f) \leqslant C_1\|f\|_{D^1}$ valid for any f (which we have already proved) we see that

$$J(f) \sim \|f\|_{D^1}.$$

Now, in order to prove that $\|g\|_{D^1} \leqslant CJ(g)$ note that because of the dyadic nature of g, we have

$$\begin{aligned}
\|g\|_{D^1} &\sim \sum_{k=0}^{\infty} 2^k \|\chi_{\{g=2^k\}}\|_{D^1} \\
&\sim \sum_{k=0}^{\infty} 2^k |\{g = 2^k\}| \log \frac{1}{|\{g = 2^k\}|} \\
&= \sum_{|\{g=2^k\}|\leqslant 2^{-2k}} 2^k |\{g = 2^k\}| \log \frac{1}{|\{g = 2^k\}|} \\
&\quad + \sum_{|\{g=2^k\}|>2^{-2k}} 2^k |\{g = 2^k\}| \log \frac{1}{|\{g = 2^k\}|} \\
&\leqslant C\left(1 + \sum_{k=0}^{\infty} 2^k \log 2^k |\{g = 2^k\}|\right) \sim \|g\|_{L(\log+L)},
\end{aligned}$$

and we have shown that $\|g\|_{(L\log+L)} \leqslant CJ(g)$, so that these inequalities prove that $\|g\|_{D^1} \leqslant CJ(g)$, finishing the proof of the lemma.

Now, to prove our theorem, take Dini function, f, on $[0, 1]$. The Dini norm of f dominates f's entropy. But the fundamental property of entropy (for a set of fixed Lebesgue measure, entropy is minimized when the set is an interval) entropy decreases as we pass from f to its nonincreasing rearrangement f^*. That is,

$$\|f\|_{D^1} \geqslant CJ(f) \geqslant CJ(f^*).$$

But f^* decreases so that, for all practical purposes, $J(f^*) = \|f^*\|_{D^1}$. Therefore $\|f\|_{D^1} \leqslant C\|f^*\|_{D^1}$, and our proof is complete.

For purposes of clarity we would like to isolate the main idea of the proof. Although it is not clear, a priori, that passing from f to f^* decreases the Dini norm, it is both a fundamental and obvious fact that this passage decreases entropy. Thus, by relating quantities like the Dini norm to the entropy, we prove our theorem.

The next application that we have in mind is of a different nature entirely. It is a problem of differentiation of the integral of a locally integrable function. More specifically, suppose that we are in R^3, and consider the two-parameter family of rectangles $R_{s,t} = \{(x, y, z) \in R^3 \mid |x| < s, |y| < t, |z| < st\}$, where $s, t > 0$. Then for which locally integrable functions $f(x, y, z)$ on R^3 do we have

$$(\sim) \quad \lim_{s,t\to 0^+} \frac{1}{|R_{st}|} \iiint_{R_{st}} f(x+\xi, y+\eta, z+\zeta)\, d\xi\, d\eta\, d\zeta = f(x, y, z)$$

for a.e. $(x, y, z) \in R^3$?

It is a consequence of the strong maximal theorem in R^3 that this differentiation of the integral works for $f \in L \log {}^+L)^2$. This result, however, is unsatisfactory because for $f \in L \log {}^+L)^2$ we already have differentiation of the integral with respect to the full family of rectangles with sides parallel to the axes. Using entropy we can find a smoothness condition on f which is not strong enough to throw f into $L \log {}^+L)^2$ and yet is sufficient to guarantee that $(\sim)$ holds a.e. The condition for $f \in L^1(Q_0)$ (Q_0 is the unit cube in R^3) is

$$\|f\| = \iint_{x,y\in Q_0} \frac{|f(x)-f(y)|}{|x-y|^3} \log \frac{1}{|x-y|}\, dx\, dy < \infty.$$

And the proof is a consequence of our earlier observations. If

$$M(f)(x) = \sup_{s,t>0} \frac{1}{|R_{st}|} \int_{R_{st}} |f(x+Z)|\, dZ$$

then the point is that when we calculate $M(\chi_Q)$, where Q is a cube, we get a function whose (we consider all f defined on Q_0) norm in weak L^1 (Q_0) is of the order of magnitude $\log(1/|Q|)\,|Q|$. If we have $f \leqslant \sum_{k,m} 2^k \chi_{Q_{k\,m}}$ then $M(f) \leqslant \sum_{k,m} 2^k M(\chi_{Q_{km}})$, so that by the Stein–Weiss theorem,

$$\|M(f)\|_{\mathrm{W}L^1} \leqslant c \sum_{k,m} 2^k\, |Q_{k,m}| \left(\log \frac{1}{|Q_{k,m}|}\right)^2.$$

If the Q_{km} are chosen efficiently then the right-hand side of the last inequality is approximately the entropy of f with respect to $\varphi(x) = x(\log(1/x))^2$, that is, we have $\| M(f)\|_{WL^1} \leqslant C\mathrm{J}_\varphi(f)$. But proceeding exactly as before, when working with the nonisotropic maximal operator M_Ω, we dominate $\mathrm{J}_\varphi(f)$ by the smoothness norm in the statement of the theorem above. Since the continuous functions are dense in the space of all functions of finite norm, we get, in the standard way, the differentiation theorem ($\sim$) above.

Before moving on, we should remark that this simple argument above provides a mathematical way of expressing the general point of view that the way to obtain results on certain maximal operators is to examine the behavior of those operators applied to a point mass. Now for some maximal functions, such as the classical Hardy–Littlewood and strong maximal functions, the general nature of the operator comes out already in the case of a point mass. For the Kakeya maximal function, the point mass case (or the case in which the function is radial) is misleading. The point is that for maximal operators where the point mass case is not general, entropy shows just how far this special case takes us toward the general function. The author confesses here for the record his belief that it will turn out when more is known that such maximal operators are rare.

Proceeding with the program, let us now consider how the familiar operators of Fourier analysis treat entropy. The Hardy–Littlewood maximal theorem may be roughly described as saying that the passage from a locally integrable function, f, to its Hardy–Littlewood maximal function, $M(f)$, preserves size of functions. The natural question to ask is "If $f \to M(f)$ preserves size of functions, does it preserve entropy?" The answer to this question is yes, in the sense that

$$E\{M(f) > \alpha\} \leqslant (C/\alpha)\ \mathrm{J}(f).$$

We also get a similar result for capacity, and our next item of business is to prove a lemma which unites both the capacitary and entropic theories.

Before stating the lemma, however, we should discuss some preliminaries. Now, basically what the lemma will say is that under certain types of small perturbations performed on a set, the entropy of the set is not increased that much. Let us define our terms. Suppose S and $\tilde{S}$ are sets, and $\tilde{S}$ is a union of intervals. Then $\tilde{S}$ is called a perturbation of S via the cover $\{I_k\}$ provided that S is covered by intervals I_k having the property that distance $(I_k, I_j) \geqslant \max(|I_k|, |I_j|)$ and such that for each k, $| S \cap I_k | = | \tilde{S} \cap I_k |$ and $\tilde{S} \cap I_k$ is a single subinterval of I_k. In other words $\tilde{S}$ and S have identical mass in each I_k, but the mass of $\tilde{S}$ in I_k is organized into an interval. $\tilde{S}$ is called a perturbation of S of it is a pertubation via some sequence $\{I_k\}$ of intervals. Our lemma can now be stated easily.

LEMMA. *Let $\mathscr{A}$ be a set function which is positive, monotone, countably subadditive, defined on the family of all subsets of the interval* $[0, 1]$, *with* $\mathscr{A}([0, 1]) = 1$.

Suppose $\mathscr{A}$ has the property that for any set $S \subset [0, 1]$ and $\tilde{S}$ a perturbation of S, we have

$$\mathscr{A}(\tilde{S}) \leqslant C\mathscr{A}(S).$$

Finally, assume that $\mathscr{A}(\bigcup I_k^\alpha) \leqslant C\alpha\mathscr{A}(\bigcup I_k)$ (for an interval I, I^α denotes the interval concentric with I only α times as large) for every sequence of intervals $\{I_k\}$ in $[0, 1]$.

Then for the Hardy–Littlewood maximal operator, M, the following estimate holds:

$$\mathfrak{A}\{M(f) > \alpha\} \leqslant \frac{A}{\alpha} \int_0^\infty \mathfrak{A}\{|f| > \alpha\}\, d\alpha.$$

Proof. Assume first that $f \geqslant 0$, and $\alpha > 0$. Then for every point $x \in \{M(f) > \alpha\}$, choose an interval I_x centered at x so that $(1/|I_x|) \int_{I_x} f(t)\, dt > \alpha$. Applying the standard covering lemma (see [8]) to the doubles, $I_x{}^2$, of the cover $\{I_x\}_{x \in \{Mf>\alpha\}}$ we have a disjoint sequence I_k such that the $(I_k{}^2)^2 = I_k{}^4$ cover $\{M(f) > \alpha\}$ and such that $(1/|I_k|) \int_{I^\Gamma} f(t)\, dt > \alpha$. Since $\mathscr{A})\bigcup I_k{}^4) \leqslant 4C \cdot \mathscr{A}(\bigcup I_k)$ and $\mathscr{A}$ is assumed monotone, we have $\mathscr{A}\{M(f) > \alpha\} \leqslant 4C\mathscr{A}(\bigcup I_k)$. So it only remains to estimate $\mathscr{A}(\bigcup I_k)$.

We may also assume, by homogeneity that $\alpha = 1$. In this case

$$\int_0^\infty \mathfrak{A}\{f > \gamma\}\, d\gamma \geqslant C_1 \sum_{k=1}^{\infty} 2^k \mathfrak{A}\left(x \in \bigcup I_k \mid 2^k \leqslant f(x) < 2^{k+1}\right).$$

Now, since the doubles of the I_k are disjoint, $\operatorname{dist}(I_k, I_j) \geqslant \max(|I_k|, |I_j|)$. Let $\{x \in U I_k \mid 2^k \leqslant f(x) < 2^{k+1}\} = S_k$. Then it is clear that because the average of f exceeds, there are perturbations $\tilde{S}_k$ of the sets S_k so that $\bigcup_{k,j}$ $(\tilde{S}k \bigcap I_j)^2\ \omega \bigcup_k I_k$. But then

$$\begin{aligned} \mathfrak{A}\left(\bigcup I_k\right) &\leqslant \mathfrak{A}\left[\bigcup_{kj} (\tilde{S}_k \cap I_j)^{2^k}\right] \leqslant \sum_k \mathfrak{A}\left[\bigcup_j (\tilde{S}_k \cap I_j)^{2^k}\right] \\ &\leqslant C \sum_k 2^k \mathfrak{A}\left[\bigcup_j (\tilde{S}_k \cap I_j)\right] = C \sum 2^k \mathfrak{A}(S_k) \sim \int_0^\infty \mathfrak{A}\{f > \gamma\}\, d\gamma, \end{aligned}$$

which is the desired estimate.

So the point here is that if I is an interval and $S_1, S_2, \ldots, S_k$ are sets contained in I with $\sum_{1 \leqslant j \leqslant k} |S_j| = |I|$, there may not be any way to translate the sets S_j by amounts τ_j so that $I = \bigcup_j (S_j + \tau_j)$. But if the S_j happen to be intervals then such translate can be found. This is the reason for discussing the action of $\mathscr{A}$ under perturbation, and this property possessed by intervals which is not shared by arbitrary sets ties the notion of entropy to the theory of the maximal function.

Next we wish to show that both entropy and capacity have the properties that we assume true of our set function $\mathscr{A}$ in the lemma above.

First, consider entropy. We shall prove that $E(\bigcup I_k^\alpha) \leqslant C\alpha E(\bigcup I_k)$ $(\alpha > 1)$ and $E(\tilde{S}) \leqslant CE(S)$ (here $\tilde{S}$ is any perturbation of S). To prove the first inequality, assume, as we may that we have only finitely many intervals, $I_1, I_2, \ldots, I_k$, and their enlargements $I_j^\alpha j \leqslant k$. Let us cover the I_j with intervals $\{J_l\}$ in such a way that (*) $\sum |J_l| \log(1/|J_l|) < E(\bigcup I_j) + \epsilon$, where ϵ is small, $\epsilon > 0$. Clearly, we may assume that the J_l are disjoint, since if $J_{l_1} \cap J_{l_2} \neq \varnothing$, we may replace J_{l_1} and J_{l_2} by their union and then (*) will be even more true. Then any I_j is contained entirely in some J_l so that $I_j^\alpha \subset J_l^\alpha$ and so $E(\bigcup I_j^\alpha) \leqslant \sum E(J_l^\alpha) \leqslant C\alpha E(J_l) < C\alpha[E(\bigcup I_j) + \epsilon]$, which ends the simple proof.

To show that $E(\tilde{S}) \leqslant CE(S)$ whenever $\tilde{S}$ is a perturbation of S, take a cover of S by intervals I_k satisfying dist $(I_j, I_k) \geqslant \max(|I_j|, |I_k|)$. Now, $\tilde{S}$ is a union of intervals $\tilde{I}_k \subset I_k$ with $|\tilde{I}_k| = |S \cap I_k|$. So let $\{J_l\}$ be a cover of S with $\sum_l |J_l| \log(1/|J_l|) < E(S) + \epsilon$. Now, the I_k naturally divide into two classes. For some k it may happen that $S \cap I_k$ is covered entirely by intervals which do not intersect any I_j for $j \neq k$. Call this set of k's K_1. If $k \notin K_1$ (then we say $k \in K_2 = {}^cK_1$) then some point $x \in S \cap I_k$ lies in some one of the J intervals, say J_k, which intersects other intervals I_j besides I_k. Thus if for each $k \in K_2$ we choose such a J_k, we have the fact that the double of J_k contains every I interval which J_k intersects. That is $\bigcup_{k\in K_2} I_k \subset \bigcup_{k\in K_2} J_k^2$. So we see that

$$E(\tilde{S}) = E\left(\bigcup \tilde{I}_k\right) \leqslant \sum_{k\in K_1} E(\tilde{I}_k) + E\left(\bigcup_{k\in K_2} I_k\right)$$

$$\leqslant \sum_{j\in K_1} \sum_{\substack{k\in K_1 \\ J_k \cap I_j \neq \phi}} E(J_k) + \sum_{k\in K_2} E(J_k^2) \leqslant \sum_{\text{all } k} E(J_k^2) < 2C(E(S) + \epsilon).$$

(The inequalities above are based on the fact that for $k \in K_1$, $|\tilde{I}_k| = |I_k \cap S|$, and because $\tilde{I}_k$ is an interval it must have smaller entropy than $I_k \cap S$ which is covered by the collection of all J_l, $l \in K_1$.) This completes the proof that entropy has the desired properties; now, on to capacity. (We shall look at C_k where

$$K(t) = 1\Big/|t| \left(\log \frac{1}{|t|}\right)^2,$$

but most of what is said generalizes immediately to a large class of kernels.)

That $C_K(\bigcup I_k^\alpha) \leqslant A\alpha C_K(\bigcup I_k)$ is shown in [5]. We shall explain here why $C_k(\tilde{S}) \leqslant AC_k(S)$ when $\tilde{S}$ is a perturbation of S. So let $\{I_k\}$ be a cover of S so that distance $(I_k, I_j) \geqslant \max(|I_k|, |I_j|)$, and let $S \cap I_k = \tilde{I}_k$. Let us put a mass distribution μ supported in $\tilde{S}$ in such a way that $K * \mu \leqslant 1$ on $\tilde{S}$, and yet $\mu(\tilde{S}) > C_k(\tilde{S}) - \epsilon$. Then if we spread an amount of mass $\mu(\mathrm{r}_k)$ uniformly over the set $S \cap I_k$ and call the resulting measure λ, then we claim that $K * \lambda \leqslant A$ on S. This would then prove that $C_k(S) \geqslant (1/A)\, C_k(\tilde{S})$, which is what we want. Now if

$$x \in I_k \cap S,\ K * \lambda(x) = \sum_{j\neq k} \int_{I_j} K(x-y)\, d\lambda(y) + \int_{I_k} K(x-y)\, d\lambda(y).$$

On I_j for $j \neq k$, $K(x - y)$ looks constant, and so $\int_{I_j} K(x - y)\, d\lambda(y) \sim \int_{I_j} K(x - y)\, d\mu(y)$ and we have assumed that $\sum \int_{I_j} K(x - y)\, d\mu(y) \leqslant 1$. The term $\int_{I_k} K(x - y)\, d\lambda(y)$ must be handled differently. We first appeal to the Hardy–Littlewood principle that $\int fg \leqslant \int f^*\, g^*$, where f and g are positive and f^* is the nonincreasing rearrangement of f. Applying this to the present situation, we have $\int_{I_k} K(x - y)\, d\lambda(y) \leqslant \int_{\tilde{I}_k} K(\bar{x} - y)\, d\tilde{\mu}(y)$ where $\bar{x}$ is the center point of $\tilde{I}_k$ and $d\tilde{\mu}$ is the uniform distribution of mass $\mu(\tilde{I}_k)$ on $\tilde{I}_k$. But since $\int_{\tilde{I}_k} K(x - y)\, d\mu(y) \leqslant 1$, $x \in \tilde{I}_k$ we have $\mu(\tilde{I}_k) \leqslant C_k(\tilde{I}_k)$, and therefore $\int_{\tilde{I}_k} K(\bar{x} - y)\, d\tilde{\mu}(y) \leqslant A$, so that finally $\int_{I_k} K(x - y)\, d\lambda(y) \leqslant A$, finishing the proof. Before looking at the situation in the right way, the inequality $E\{M(f) > \alpha\} \leqslant (C/\alpha)\, J(f)$ may lead us to expect a similar inequality for singular integral operators, since it is usually true that, except for questions relating directly to the positivity of the operator M, what is true for the maximal function is true for singular integrals. But in the case of entropy these operators are genuinely different. The reason seems to be intimately connected to the fact that entropy fails to be dilation invariant, and to the unsatisfactory L^2 theory of entropy. For the maximal operator the space L^∞ is the most basic, and for singular integrals the space L^2 is. But there is no Plancheral theorem for entropy, and yet it is obvious that bounded functions have finite entropy.

We shall first show that the inequality $E\{|H(f)| > \alpha\} \leqslant (C/\alpha)\, J(f)$ cannot hold, where H is the Hilbert transform of a function f on the circle. After the counterexample, we shall show how to apply Calderón–Zygmund theory to singular integrals, by considering the Dini class instead of all functions of finite entropy, and this will put to use much of the work done in the previous sections.

Now, suppose we take for our function f, the characteristic function of the set $S_{N\delta}$ (for a discussion of this set, see Section 2), and fix $\alpha > 0$ very large. Then $H(f)$ is the dilate via the dilation $x \to nx$, $n \sim 1/N\delta$ of the function $\log(\delta/|x - \delta|)$. Hence, the set $\{|H(f)| > \alpha\}$ will be essentially $S_{Ne^\alpha, \delta e^{-\alpha}}$, and as long as we choose N large, and then take δ so small that $\delta e^{-\alpha} < e^{-Ne^\alpha}$ we shall have $E\{|H(f)| > \alpha\} \sim 1$, and since all f under consideration have uniformly bounded entropies, this clearly rules out an inequality like $E\{|H(f)| > \alpha\} \leqslant (C/\alpha)\, J(f)$.

Nevertheless one can use the methods we have developed to carry through the program of Calderón–Zygmund for treating singular integral operators, provided we are willing to consider smoothness classes, rather than classes of functions with finite entropy. We shall work in R^n, and consider kernels, $\Omega(x)$, which are nomogeneous of degree 0 on R^n and satisfy $\int_{S^{n-1}} \Omega(x)\, d\sigma(x) = 0$, as well as $\int_0^1 \omega_\infty(\delta)(d\delta/\delta) < \infty$, where

$$\omega_\infty(\delta) = \sup_{\substack{x, y \in S^{n-1} \\ |x - y| \leqslant \delta}} |\Omega(x) - \Omega(y)|.$$

What we will now obtain is a decomposition into good and bad parts of a function $f \in D^1$. We need one more definition. A function f defined on R^n and supported in the unit cube, Q_0, will be said to satisfy an L^2 Dini condition provided

$$\|f\|^2_{L^2(Q_0)} + \iint_{x,y\in Q_0} \frac{|f(x)-f(y)|^2}{|x-y|^n}\,dx\,dy = \|f\|^2_{D^2} < \infty.$$

We then have the following theorem:

THE CALDERÓN–ZYGMUND DECOMPOSITION FOR ENTROPY. *Let f be a function defined on R^n with support in Q_0 with $f \in D^1$. Let Ω be as above, and set*

$$T_\epsilon(f)(x) = \int_{|y|>\epsilon} f(x-y)\,\frac{\Omega(y)}{|y|^n}\,dy,\ \epsilon > 0, \quad \text{and} \quad T(f)(x) = \lim_{\epsilon\to 0} T_\epsilon(f)(x).$$

If $\alpha > 0$, there is $F \in D^2$ with $\|F\|^2_{D^2} \leqslant \alpha\,\|f\|_{D^1}$ so that

$$E\{x\in Q_0 \mid |T(f)(x) - F(x)| > \alpha\} \leqslant (C/\alpha)\,\|f\|_{D^1}.$$

COROLLARY. *With respect to the appropriate kernel, K (namely*

$$K(x) = 1\Big/|x|\left(\log\frac{1}{|x|}\right)^2,$$

we have the following capacitary estimate for the operator T:

$$C_k\{x\in Q_0 \mid |T(f)(x)| > \alpha\} \leqslant (C/\alpha)\,\|f\|_{D^1}.$$

The corollary is an immediate consequence of Carleson's maximal theorem for capacity: $C_k\{M(f) > \alpha\} \leqslant (C/\alpha^2)\,\|f\|^2_{D^2}$ and the Calderón–Zygmund decomposition for entropy, since

$$\begin{aligned} C_k\{x\in Q_0 \mid |T(f)(x)| > 2\alpha\} &\leqslant E\{x\in Q_0 \mid |T(f)(x) - F(x)| > \alpha\} \\ &\quad + C_k\{x\in Q_0 \mid |F(x)| > \alpha\} \\ &\leqslant \frac{C}{\alpha}\,\|f\|_{D^1} + \frac{C}{\alpha^2}\,C\alpha\,\|f\|_{D^1} \sim \frac{\|f\|_{D^1}}{\alpha}. \end{aligned}$$

Proceeding to the proof of the decomposition theorem, we take $f \in D^1$, and $\alpha > 0$. We may assume $f \geqslant 0$ and $\alpha > 0$. By the decomposition of functions in L^1, we find subcubes $\{Q_k\}$, $k \geqslant 1$, of the unit cube in R^n, Q_0, so that

$$\alpha < \frac{1}{|Q_k|}\int_{Q_k} f(x)\,dx \leqslant 2^n\alpha, \text{ and } f(x) \leqslant \alpha \text{ outside } \bigcup_k Q_k.$$

We cannot use the cubes Q_k directly, but there will be other cubes Q_k' which will serve instead. They are determined as follows: Since for each $x \in \bigcup_{k \geqslant 1} Q_k$, $M(f)(x) > \alpha$, we have, by the Maximal theorem for entropy, that

$$E\left(\bigcup_k Q_k\right) \leqslant \frac{C}{\alpha} J(f) \leqslant \frac{C}{\alpha} \|f\|_{D^1}.$$

Because of this, there are cubes Q'_k so that $\bigcup Q_k \supset \bigcup Q_k$ and

$$\sum |Q'_k| \log \frac{1}{|Q'_k|} \leqslant \frac{C}{\alpha} \|f\|_{D^1}.$$

Now, let us decompose the function f as follows: $f = g + h + b$, where $g(x) = f(x)$, where $f \leqslant \alpha$, and $g(x) = \alpha$ otherwise. Then we let

$$h(x) = \sum_{k \geqslant 1} \left(\frac{1}{|Q'_k|} \int_{Q'_k} f(y)\, dy - \frac{1}{|Q'_k|} \int_{Q'_k} g(y)\, dy\right) \chi_{Q'_k}(x)$$

and finally we set $b = f - g - h$. Then since g is merely a truncation of f at height α, g is smoother than f so that $\|g\|_{D^1} \leqslant \|f\|_{D^1}$. On the other hand g is bounded by α so that $\|g\|_{D^2}^2 \leqslant \alpha \|g\|_{D^1} \leqslant \alpha \|f\|_{D^1}$. What about h?

We may assume that the Q_k in the decomposition are dyadic, and the same for the cubes Q'_k covering the Q_k. Then the point is that

$$\frac{1}{|Q'_k|} \int_{Q'_k} f(y)\, dy = \frac{|Q'_k - \bigcup_j Q_j|}{|Q'_k|} \left(\frac{1}{|Q'_k - \bigcup_j Q_j|} \int_{Q'_k - \bigcup Q_j} f(y)\, dy\right)$$

$$+ \sum_j \frac{|Q_j|}{|Q'_k|} \left(\frac{1}{|Q_j|} \int_{Q_j} f(y)\, dy\right).$$

Since $f \leqslant \alpha$ on $Q'_k - \bigcup_j Q_j$ and $(1/|Q_j|) \int_{Q_j} f(y)\, dy \sim \alpha$ we have

$$\frac{1}{|Q'_k|} \int_{Q'_k} f(y)\, dy \leqslant \frac{|Q'_k - \bigcup Q_j|}{|Q'_k|} \alpha + \sum_j \frac{|Q_j|}{|Q'_k|} C\alpha \leqslant C'\alpha.$$

This tells us that $|h| \leqslant C'\alpha$. Therefore

$$\|h\|_{D^1} \leqslant C'\alpha \sum_k^k \|\chi_{Q'_k}\|_{D^1} \leqslant C'\alpha \sum |Q'_k| \log \frac{1}{|Q'_k|}$$

$$\leqslant C''\alpha \frac{1}{\alpha} \|f\|_{D^1} = C'' \|f\|_{D^1}.$$

As before, since $|h| \leqslant \alpha$, $\|h\|_{D^2} \leqslant C''\alpha \|f\|_{D^1}$. But now, let $F = T(g + h)$. We know that the operator T is bounded on L^2 and commutes with translations, so that it is bounded on D^2. Therefore $\|F\|_{D^2} \leqslant C \|g + h\|_{D^2} \leqslant C \|f\|_{D^1}\alpha$, as advertised. (Here the letter C indicates an absolute constant which may not be the same absolute constant if we use it in two different expressions.)

And now in order to treat $T(b)$, we consider the doubles of the cubes Q_k', $\tilde{Q}_k'$ with centers y_k, and let $x \notin \tilde{Q}_k'$ $(x \in Q_0)$. Then

$$| T(b)(x)| = \left| \sum_k \int_{Q_k'} b(y)[K(x-y) - K(x-y_k)]\, dy \right|$$

$$\leqslant \sum_k \int_{Q_k'} | b(y)| \, | K(x-y) - K(x-y_k)| \, dy.$$

Since entropy is equivalent to a norm, we have

$$\mathrm{J}(\chi_{Q_0 - \cup \tilde{Q}_j'} \, | \, T(b)|) \leqslant C \sum_k \int_{Q_k} | b(y)| \, \mathrm{J}_x(| K(x-y) - K(x-y_k)| \, \chi_{Q_0 - \cup \tilde{Q}_j'})\, dy.$$

(Here $\mathrm{J}_x(f(x, y))$ means that for a fixed y we consider the entropy of the function $x \to f(x, y)$.)

Let us now estimate

$$J_x(| K(x-y) - K(x-y_k)| \, \chi_{Q_0 - \tilde{Q}_k}) \, | K(x-y) - K(x-y_k)|$$

$$\leqslant \frac{| \Omega(x-y_k) - \Omega(x-y)|}{| x-y_k |^n} + | \Omega(x-y)| \left| \frac{1}{| x-y_k |^n} - \frac{1}{| x-y |^n} \right|$$

$$\leqslant C \left[\frac{\omega_\infty(| y-y_k |/| x-y_k |)}{| x-y_k |^n} + \| \Omega \|_\infty \, | y-y_k | \, \frac{1}{| x-y_k |^{n+1}} \right],$$

$$\mathrm{J}_x(| K(x-y) - K(x-y_k)| \, \chi_{Q_0 - \tilde{Q}_k'})$$

$$\leqslant C \left[\sum_{m=0}^\infty \frac{\omega_\infty(| y-y_k |/2^m\delta)}{2^{mn}\delta^n} (2^m\delta)^n \log \frac{1}{(2^m\delta)} \right.$$
$$\left. + \| \Omega \|_\infty \, | y-y_k | \sum_{m=0}^\infty \frac{1}{(2^m\delta)^{n+1}} (2^m\delta)^n \log \frac{1}{(2^m\delta)} \right]$$

(here $\delta = \operatorname{diam}(Q)_k'$)

$$\leqslant C \left[\sum_m \omega_\infty \left(\frac{1}{2^m} \right) \log \frac{1}{| Q_k' |} + \log \frac{1}{| Q_k' |} \right].$$

Since

$$\sum \omega_\infty \left(\frac{1}{2^m} \right) \sim \int_0^1 \frac{\omega_\infty(\delta)}{\delta}\, d\delta,$$

we have

$$\mathrm{J}_x(|K(x-y)-K(x-y_k)|\,\chi_{Q_0-\cup_j\tilde{Q}_j'}) \leqslant C \log \frac{1}{|Q_k'|}$$

so that

$$\mathrm{J}(\chi_{Q_0-\cup\tilde{Q}_j'}\,|T(b)|) \leqslant C \sum_k \left(\int_{Q_k'} |b(y)|\,dy\right) \log \frac{1}{|Q_k'|}$$

$$\leqslant C \sum_k (\alpha\,|Q_k'|) \log \frac{1}{|Q_k'|} \leqslant C\alpha \left(\frac{\|f\|_{D^1}}{\alpha}\right) = C\,\|f\|_{D^1}.$$

It follows that $E\{|T(b)| > \alpha\} \leqslant (C/\alpha)\,\|f\|_{D^1}$, finishing our proof.

Acknowledgment

I would like to express my deep gratitude to my adviser, E. M. Stein, whose enormous enthusiasm and inspiration were invaluable to me throughout my work.

References

1. D. R. Adams, Maximal operators and capacity, *Proc. Amer. Math. Soc.* **34** (1972), 152–156.
2. A. Calderón and A. Zygmund, On the existence of certain singular integrals, *Acta Math.* **88** (1952), 85–139.
3. A. Calderón and A. Zygmund, On singular integrals, *Amer. J. Math.* **78** (1956), 289–309.
4. C. Calderón, Differentiation through starlike sets in R^n, *Studia Math.* **48** (1973), 1–13.
5. L. Carleson, "Selected Problems in Exceptional Sets," Van Nostrand, Princeton, N.J., 1967.
6. H. Federer, "Geometric Measure Theory," Springer-Verlag, New Yo-k, 3969.
7. A. Garsia, On the smoothness of functions satisfying certain integral inequalities, *in* "Functional Analysis," (C. O. Wilde, Ed.), Academic Press, New York, 1970.
8. E. M. Stein, "The Analogues of Fatou's Theorem and Estimates for Maximal Functions," C.I.M.E., 1967.
9. E. M. Stein, "Singular Integrals and Differentiability Properties of Functions," Princeton Univ. Press, Princeton, N.J., 1970.
10. E. M. Stein and N. J. Weiss, On the convergence of Poisson integrals, *Trans. Amer. Math. Soc.* **140** (1969), 35–54.
11. S. J. Taylor, On the connexion between Hausdorff measures and generalized capacity, *Proc. Cambridge Philos. Soc.* **57** (1961), 524–531.
12. M. Weiss, A. Calderón, and A. Zygmund, On the existence of singular integrals, *Proc. Symp. Pure Math.* **10** (1967), 56–73.
13. A. Zygmund, "Trigonometric Series," Cambridge Univ. Press, London/New York, 1968.